

Buenas prácticas en análisis de datos en investigación

Lima, Perú

Dr. José Ventura-León

Marzo 8, 2022

ID	EDAD	SEXO	GRADO DE INSTRUCCIÓN	PROCEDENCIA	IDA1	IDA2	IDA3	IDA4	IDA5	IDA6	IDA7	IDA8	IDA9	IDA10	IDA11	IDA12	IDA13	IDA14	IDA15	IDA16	IDA17	IDA18	IDA19	IDA20	EMOCIONALIDAD	PREOCUPACIÓN	IPPPA1	IPPPA2	IPPPA3
1	10	1	2	1	2	1	2	1	1	4	3	4	3	4	2	4	2	1	1	4	4	3		=SUMA(G2:H2;I2;K2;L2;M2;N2;P2;R2;T2;S2)		3	1	5	
2	9	1	2	1	2	1	2	1	4	1	SUMA(número1; [número2]; [número3]; [número4]; [número5]; [número6]; [número7]; [número8]; [número9]; [número10]; [número11]; [número12];																		
3	9	1	2	1	1	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	44	28	4	1	5
4	10	1	2	1	4	2	2	1	3	1	3	2	3	1	1	4	2	2	1	1	2	1	1	2	20	12	3	3	5
5	9	1	2	1	2	2	3	1	3	2	1	2	2	3	4	3	1	1	2	2	3	2	3	4	21	20	5	1	5
6	10	1	2	1	2	4	4	4	3	4	4	4	4	3	4	4	4	4	3	4	3	4	4	4	42	27	4	2	3
7	9	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	21	14	5	2	5
8	9	1	2	1	2	4	4	2	4	4	4	4	3	4	2	2	4	2	1	3	4	3	1	3	34	20	4	2	3
9	9	1	2	1	4	2	3	2	1	2	3	4	3	3	3	3	3	4	3	3	4	3	3	2	32	21	5	3	2
10	9	1	2	1	4	1	1	1	4	1	4	4	4	1	4	1	1	2	1	4	1	1	4	1	24	13	3	1	4
11	10	1	2	1	4	1	1	1	1	1	2	2	2	3	2	2	1	1	1	4	3	1	1	1	15	15	4	3	3
12	9	1	2	1	2	3	4	1	3	2	2	1	2	3	1	4	2	1	1	3	4	3	2	1	20	20	4	2	5
13	10	1	2	1	2	3	4	2	2	2	4	4	4	4	4	4	4	2	4	2	4	4	4	4	35	26	3	4	2
14	9	1	2	1	3	2	2	1	3	1	1	1	1	2	1	2	1	1	1	1	1	2	2	1	13	11	1	3	5
15	9	1	2	1	2	4	4	2	4	3	3	3	2	3	4	2	1	2	2	3	1	2	3	30	17	4	2	5	
16	9	2	2	1	4	3	4	3	2	1	3	2	2	4	3	3	2	2	2	3	3	2	1	1	27	17	4	3	5
17	10	2	2	1	1	4	4	1	1	4	4	2	3	4	2	4	2	3	4	2	4	4	4	1	33	23	5	3	5
18	9	2	2	1	2	4	3	4	3	4	4	4	2	4	3	4	2	1	4	1	3	4	3	1	35	20	5	3	1
19	9	2	2	1	2	4	2	4	1	4	4	3	2	4	2	3	4	1	4	2	3	1	3	1	34	17	3	1	5
20	9	2	2	1	3	1	2	1	2	1	2	2	2	4	2	4	1	1	2	2	3	4	1	3	17	21	3	3	3
21	10	2	2	1	4	1	1	1	1	1	1	1	2	1	1	4	1	1	1	1	1	1	2	1	12	11	5	2	5
22	10	2	2	1	3	1	1	1	4	1	1	1	1	2	2	3	3	1	1	3	3	4	1	1	14	17	5	3	5
23	9	2	2	1	2	1	3	1	1	1	2	2	2	4	2	1	1	1	2	1	4	2	1	1	18	14	4	5	5
24	9	2	2	1	4	2	2	1	1	1	2	3	2	2	1	2	1	3	1	4	2	3	4	2	19	19	3	2	3

- Un procedimiento clásico, es realizar sumas a los ítems

Sin embargo este proceso no considera varios supuestos

Supuestos básicos cuando se trabaja con instrumentos autoinformados

ID	EDAD	SEXO	GRADO DE INSTRUCCIÓN	PROCEDENCIA	IDA1	IDA2	IDA3	IDA4	IDA5	IDA6	IDA7	IDA8	IDA9	IDA10	IDA11	IDA12	IDA13	IDA14	IDA15	IDA16	IDA17	IDA18	IDA19	IDA20	EMOCIONALIDAD	PREOCUPACIÓN	IPPPA1	IPPPA2	IPPPA	
1	10	1	2	1	2	1	2	1	1	4	3	4	3	4	2	4	2	1	4	4	4	3	4	3	=SUMA(G2;H2;I2;K2;L2;M2;N2;P2;R2;T2;S2)			3	1	5
2	9	1	2	1	2	1	2	1	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	44	28	4	1	5	
3	9	1	2	1	1	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	20	12	3	3	5	
4	10	1	2	1	4	2	2	1	3	1	3	2	3	1	1	4	2	2	1	1	2	1	1	2	21	20	5	1	5	
5	9	1	2	1	2	2	3	1	3	2	1	2	2	3	4	3	1	1	2	2	3	2	3	4	42	27	4	2	3	
6	10	1	2	1	2	4	4	4	4	3	4	4	4	4	3	4	4	4	4	3	4	4	4	4	21	14	5	2	5	
7	9	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	34	20	4	2	3	
8	9	1	2	1	2	4	4	2	4	4	4	4	3	4	2	2	4	2	1	3	4	3	1	3	32	21	5	3	2	
9	9	1	2	1	4	2	3	2	1	2	3	4	3	3	3	3	4	3	3	4	3	3	2	24	13	3	1	4		
10	9	1	2	1	4	1	1	1	4	1	4	4	4	1	4	1	2	1	2	1	4	1	1	4	15	15	4	3	3	
11	10	1	2	1	4	1	1	1	1	2	2	2	2	3	2	2	1	1	1	4	3	1	1	1	20	20	4	2	5	
12	9	1	2	1	2	3	4	1	3	2	2	1	2	3	1	4	2	1	1	3	4	3	2	1	35	26	3	4	2	
13	10	1	2	1	2	3	4	2	2	2	2	4	4	4	4	4	2	4	2	4	4	4	4	4	13	11	1	3	5	
14	9	1	2	1	3	2	2	1	3	1	1	1	1	2	1	2	1	1	1	1	1	2	2	1	30	17	4	2	5	
15	9	1	2	1	2	4	4	2	4	3	3	3	3	2	3	4	2	1	2	2	3	1	2	3	27	17	4	3	5	
16	9	2	2	1	4	3	4	3	2	1	3	2	2	4	3	3	2	2	2	3	3	2	1	1	33	23	5	3	5	
17	10	2	2	1	1	4	4	1	1	4	4	2	3	4	2	4	2	3	4	2	4	4	4	1	35	20	5	3	1	
18	9	2	2	1	2	4	3	4	3	4	4	4	2	4	3	4	2	1	4	1	3	4	3	1	34	17	3	1	5	
19	9	2	2	1	2	4	2	4	1	4	4	3	2	4	2	3	4	1	4	2	3	1	3	1	17	21	3	3	3	
20	9	2	2	1	3	1	2	1	2	1	2	2	2	4	2	4	1	1	2	2	3	4	1	3	12	11	5	2	5	
21	10	2	2	1	4	1	1	1	1	1	1	2	1	1	1	4	1	1	1	1	1	1	2	1	14	17	5	3	5	
22	10	2	2	1	3	1	1	1	4	1	1	1	1	2	2	3	3	1	1	3	3	4	1	1	18	14	4	5	5	
23	9	2	2	1	2	1	3	1	1	1	2	2	2	4	2	1	1	1	2	1	4	2	1	1	19	19	3	2	3	
24	9	2	2	1	4	2	2	1	1	1	2	3	2	2	1	2	1	3	1	4	2	3	4	2						

1. Las variables observadas tiene naturaleza tipo Likert.
2. Asume igualdad de varianza entre los ítems (tau-equivalencia)

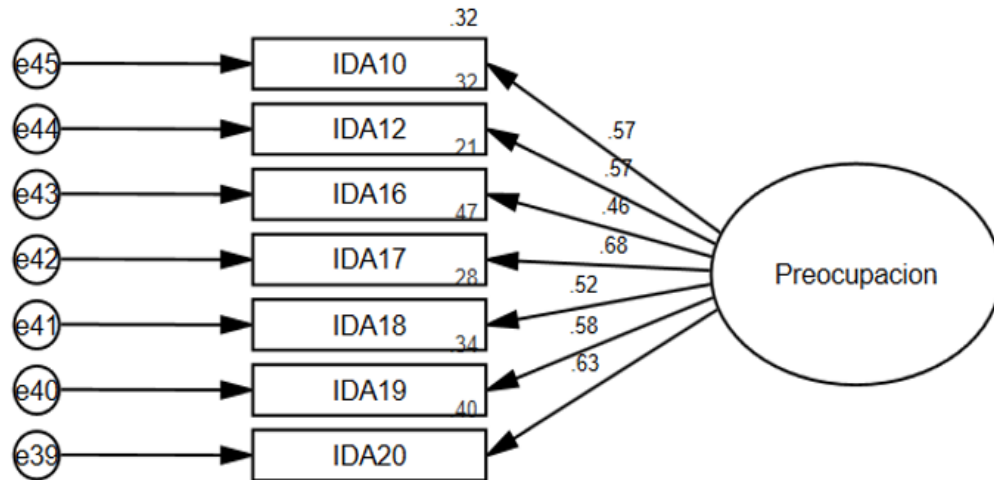
Supuestos básicos cuando se trabaja con instrumentos autoinformados

ID	EDAD	SEXO	GRADO DE INSTRUCCIÓN	PROCEDENCIA	IDA1	IDA2	IDA3	IDA4	IDA5	IDA6	IDA7	IDA8	IDA9	IDA10	IDA11	IDA12	IDA13	IDA14	IDA15	IDA16	IDA17	IDA18	IDA19	IDA20	EMOCIONALIDAD	PREOCUPACIÓN	IPPPA1	IPPPA2	IPPPA
1	10	1	2	1	2	1	2	1	1	4	3	4	3	4	2	4	2	1	1	4	4	3		=SUMA(G2;H2;I2;K2;L2;M2;N2;P2;R2;T2;S2)			3	1	5
2	9	1	2	1	2	1	2	1	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	44	28	4	1	5
3	9	1	2	1	1	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	20	12	3	3	5
4	10	1	2	1	4	2	2	1	3	1	3	2	3	1	1	4	2	2	1	1	2	1	1	2	21	20	5	1	5
5	9	1	2	1	2	2	3	1	3	2	1	2	2	3	4	3	1	1	2	2	3	2	3	4	42	27	4	2	3
6	10	1	2	1	2	4	4	4	4	3	4	4	4	4	3	4	4	4	4	3	4	4	4	4	21	14	5	2	5
7	9	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	34	20	4	2	3
8	9	1	2	1	2	4	4	2	4	4	4	3	4	2	2	4	2	1	3	4	3	1	3	3	32	21	5	3	2
9	9	1	2	1	4	2	3	2	1	2	3	4	3	3	3	3	3	4	3	3	4	3	3	2	24	13	3	1	4
10	9	1	2	1	4	1	1	1	4	1	4	4	4	1	4	1	1	2	1	4	1	1	4	1	15	15	4	3	3
11	10	1	2	1	4	1	1	1	1	1	2	2	2	3	2	2	1	1	1	4	3	1	1	1	20	20	4	2	5
12	9	1	2	1	2	3	4	1	3	2	2	1	2	3	1	4	2	1	1	3	4	3	2	1	35	26	3	4	2
13	10	1	2	1	2	3	4	2	2	2	2	4	4	4	4	4	2	4	2	4	4	4	4	4	13	11	1	3	5
14	9	1	2	1	3	2	2	1	3	1	1	1	1	2	1	2	1	1	1	1	1	2	2	1	30	17	4	2	5
15	9	1	2	1	2	4	4	2	4	3	3	3	3	2	3	4	2	1	2	2	3	1	2	3	27	17	4	3	5
16	9	2	2	1	4	3	4	3	2	1	3	2	2	4	3	3	2	2	3	3	2	1	1	1	33	23	5	3	5
17	10	2	2	1	1	4	4	1	1	4	4	2	3	4	2	4	2	3	4	2	4	4	4	1	35	20	5	3	1
18	9	2	2	1	2	4	3	4	3	4	4	4	2	4	3	4	2	1	4	1	3	4	3	1	34	17	3	1	5
19	9	2	2	1	2	4	2	4	1	4	4	3	2	4	2	3	4	1	4	2	3	1	3	1	17	21	3	3	3
20	9	2	2	1	3	1	2	1	2	1	2	2	2	4	2	4	1	1	2	2	3	4	1	3	12	11	5	2	5
21	10	2	2	1	4	1	1	1	1	1	1	2	1	1	1	4	1	1	1	1	1	1	2	1	14	17	5	3	5
22	10	2	2	1	3	1	1	1	4	1	1	1	1	2	2	3	3	1	1	3	3	4	1	1	18	14	4	5	5
23	9	2	2	1	2	1	3	1	1	1	2	2	2	4	2	1	1	1	2	1	4	2	1	1	19	19	3	2	3
24	9	2	2	1	4	2	2	1	1	1	2	3	2	2	1	2	1	3	1	4	2	3	4	2					

3. Se obtiene una variable no observable mediante la sumatoria de ítems. Es decir es un procedimiento aritmético

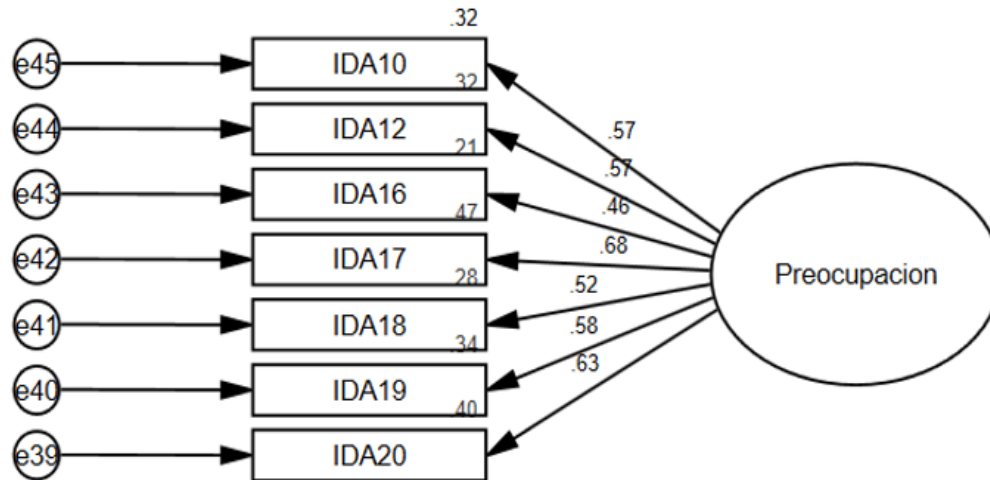
4.No incluye el error de medida de cada indicador $[X = V+e]$

Pero que sucede en la realidad...



1. Las variables observadas son casi siempre de naturaleza ordinal.
2. Desigualdad de varianza (modelos cogenéricos)

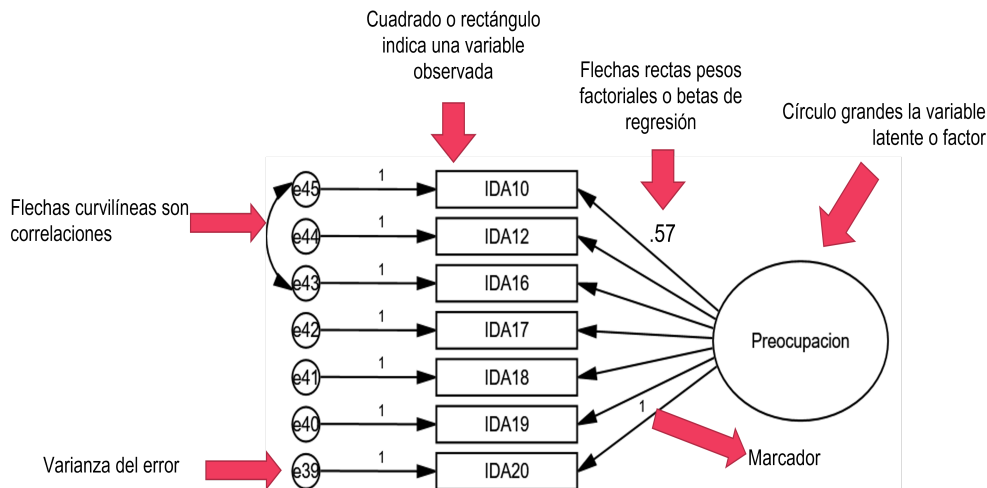
Pero que sucede en la realidad...



3. Las variables no observadas deben realizarse mediante procedimientos algebraicos

4. Toda medida tiene un error de medición.

Importante conocer los elementos de SEM



Índices de ajuste	Nivel de aceptación
Chi-cuadrado	$p > 0.05$
Chi-cuadrado entre grados de libertad (χ^2/gl)	< 2 (Tabachnik y Fidell, 2007) < 5 (Wheaton, Muthen, Alwin, & Summers, 1977)
RMSEA (Root Mean Square Error of Approximation)	0.01 (excelente), 0.05 (bueno) y 0.08 (mediocre) (MacCallum, Browne, & Sugawara, 1996)
SRMR (Standardized Root Mean Square Residual)	$< .08$ (Hu y Bentler, 1999)
CFI (Comparative Fit Index)	> 0.95

jamovi - base de datos

Analyses

Exploration T-Tests ANOVA Regression Frecuencias Factor Base R MAJOR medmod MetaModel TOSTER Walrus Modules

Confirmatory Factor Analysis

MODEL FIT

Test for Exact Fit

χ^2 test

Fit Measures

CFI
 TLI
 SRMR
 RMSEA
 AIC
 BIC

Additional Output

Post-Hoc Model Performance

Residuals observed correlation matrix
 Highlight values above

Modification indices
 Highlight values above

Plots

Path diagram

Confirmatory Factor Analysis

Factor Loadings

Factor	Indicator	Estimate	SE	Z	p
Factor 1	IDA10	0.634	0.0779	8.14	< .001
	IDA12	0.689	0.0839	8.21	< .001
	IDA16	0.512	0.0796	6.43	< .001
	IDA17	0.759	0.0743	10.22	< .001
	IDA18	0.588	0.0788	7.46	< .001
	IDA19	0.642	0.0768	8.36	< .001
	IDA20	0.715	0.0779	9.17	< .001

Factor Estimates

Factor Covariances

	Estimate	SE	Z	p
Factor 1 Factor 1	1.00*			

* fixed parameter

Model Fit


```

1 library(haven)
2 Base_de_Datos_Final_Hugo <- read_sav("C:/Users/admin/Desktop/Asesorados de Tesis/Hugo Venegas/Anal
3 View(Base_de_Datos_Final_Hugo)
4 labels(Base_de_Datos_Final_Hugo)
5
6 Padres <- subset(Base_de_Datos_Final_Hugo,
7   select = c("IDA10", "IDA12", "IDA16", "IDA17", "IDA18", "IDA19", "IDA20",
8     "IDA2", "IDA3", "IDA4", "IDA6", "IDA7", "IDA8", "IDA9", "IDA11", "IDA13"
9     "IPPPA1", "IPPPA3", "IPPPA5", "IPPPA7", "IPPPA9", "IPPPA11", "IPPPA13",
10    "IPPPA18", "IPPPA16", "IPPPA14", "IPPPA12", "IPPPA10", "IPPPA8", "IPPPA6
11    "IPPPA1", "IPPPA3", "IPPPA5", "IPPPA7", "IPPPA9", "IPPPA11", "IPPPA13",
12    "IPPPA18", "IPPPA16", "IPPPA14", "IPPPA12", "IPPPA10", "IPPPA8", "IPPPA6
13
14
15 ###Modelo 1 - padre### NA es liberar el parametro para que el error de preocupacion y emocionalida
16 model12 <- 'preocupacion =~ IDA10 + IDA12 + IDA16 + IDA17 + IDA18 + IDA19 + IDA20
17   emocionalidad =~ IDA2 + IDA3 + IDA4 + IDA6 + IDA7 + IDA8 + IDA9 + IDA11 + IDA13
18 <

```

Console C:/Users/admin/Desktop/Asesorados de Tesis/Hugo Venegas/Analisis de datos/R/

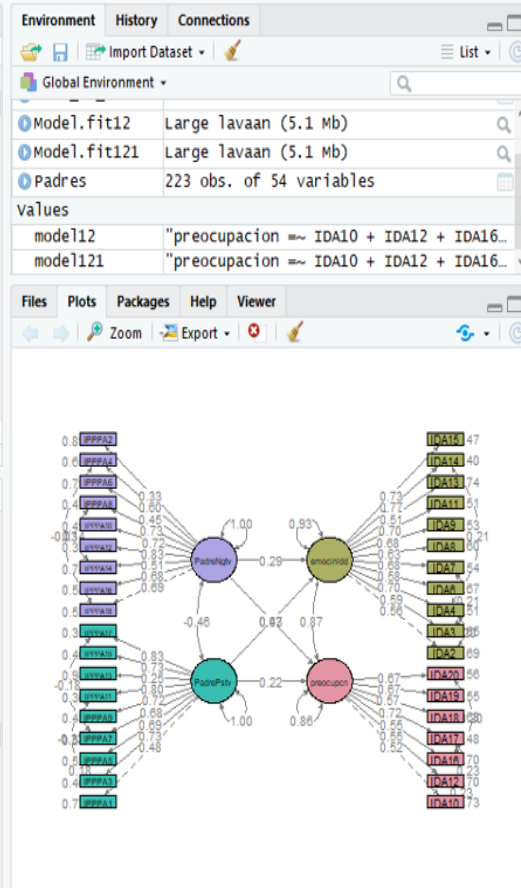
Estimator	DWLS	Robust
Model Fit Test Statistic	758.412	166.186
Degrees of freedom	577	92
P-value (Chi-square)	0.000	0.000
Scaling correction factor		4.564
for the mean and variance adjusted correction (WLSMV)		

Model test baseline model:

	11332.494	1192.561
Minimum Function Test Statistic		
Degrees of freedom	630	66
P-value	0.000	0.000

User model versus baseline model:

	0.983	0.934
Comparative Fit Index (CFI)		
Tucker-Lewis Index (TLI)	0.981	0.953



Data y organización

```
#Cargando las librerías a utilizarse
pacman::p_load(qgraph, tidyverse, correlation)

#cargando la data, que es de la librería qgraph
data(big5)
big5_new <- big5 %>%
  as_tibble() %>%
  select(starts_with("N"),starts_with("E")) %>%
  select(N1:N46,E2:E42) %>%
  rowwise() %>%
  mutate(
    Neuroticismo = sum(c_across(c(N1:N46))),
    Extraversión = sum(c_across(c(E2:E42)))
  )
```

Correlaciones clásicas (Pearson)

```
##Cálculo de correlación de Pearson
correlation_pearson <- big5_new %>%
  select(Neuroticismo:Extraversión) %>%
  correlation(method = "pearson",
             p_adjust = "none") %>% as_tibble() %>% dplyr::select(Parameter1,
             Parameter2, r, CI, CI_low, CI_high) %>%
  mutate(across(where(is.numeric), round, 2))
knitr::kable(correlation_pearson, format = "html", booktabs = TRUE, \
```

Parameter1	Parameter2	r	CI	CI_low	CI_high
Neuroticismo	Extraversión	-0.26	0.95	-0.34	-0.18

Correlaciones clásicas (Spearman)

```
#Cálculo de correlación de Spearman
correlation_spearman <- big5_new %>%
  select(Neuroticismo:Extraversión) %>%
  correlation(method = "spearman",
             p_adjust = "none") %>% as_tibble() %>% dplyr::select(Parameter1,
             mutate(across(where(is.numeric), round, 2))
knitr::kable(correlation_spearman, format = "html", booktabs = TRUE,
```

Parameter1	Parameter2	rho	CI	CI_low	CI_high
Neuroticismo	Extraversión	-0.25	0.95	-0.33	-0.16

Modelo de ecuaciones estructurales

```
library(lavaan)
```

```
## This is lavaan 0.6-10  
## lavaan is FREE software! Please report any bugs.
```

```
# 1. Especificación del modelo
```

```
model <- "Neuro =~ N1+N6+N11+N16+N21+N31+N36+N41+N46  
         Extra =~ E2+E7+E12+E17+E22+E27+E32+E37+E42"
```

```
# 2. Estimación del modelo
```

```
fit = sem(model,  
          data = big5,  
          estimator = "WLSMV",  
          mimic = "Mplus",  
          ordered = TRUE)
```

Output de correlación latente

```
SEM <- lavaan::inspect(fit, what = "std")$psi %>% as_tibble() %>%  
  mutate(across(where(is.numeric), round, 2))  
  
knitr::kable(SEM)
```

Neuro	Extra
1.00	-0.81
-0.81	1.00

bondades de ajuste

```
bondad <-  
  lavaan::fitMeasures(fit, c("chisq.scaled", "df.scaled", "srmr", "wrmr",  
  knitr::kable(bondad, format = "html", caption = "Bondades de ajuste")
```

	value
chisq.scaled	745.40
df.scaled	134.00
srmr	0.09
wrmr	1.73
cfi.scaled	0.76
tli.scaled	0.72
rmsea.scaled	0.10

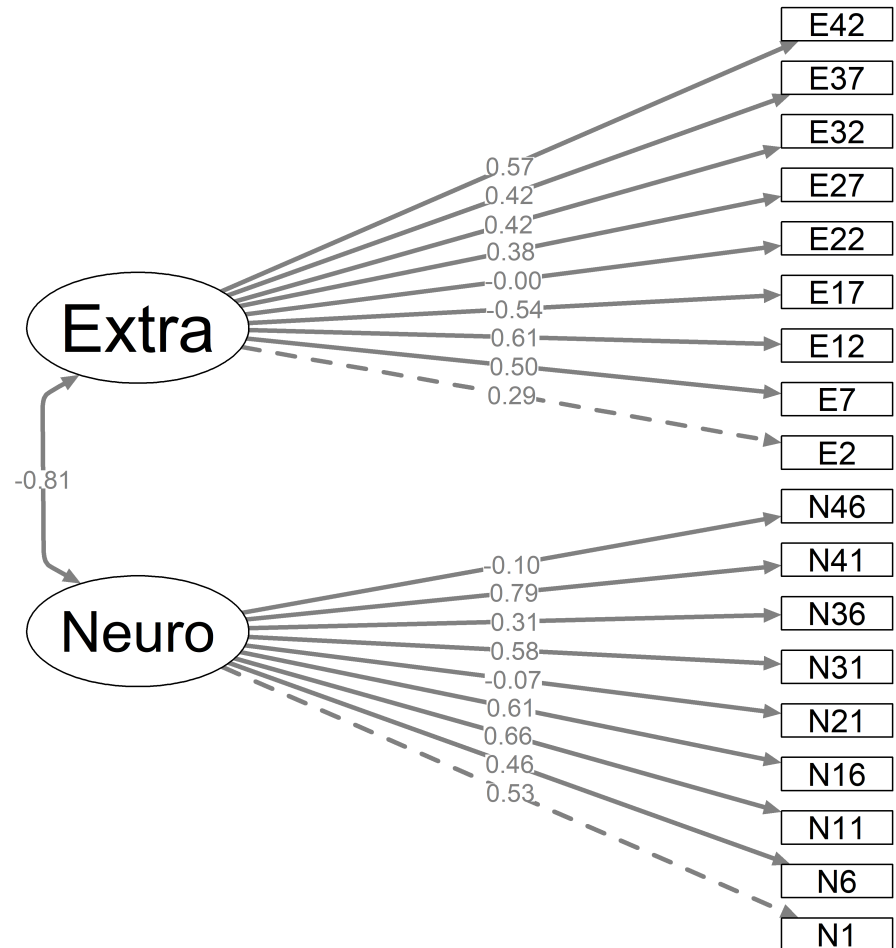
Reespecificación del modelo

```
mod <- modificationindices(fit, sort. = T, minimum.value = 20)%>%  
  select(lhs:epc) %>% mutate(across(where(is.numeric), round, 2)) %>%  
  filter(op == "~~")  
knitr::kable(mod, format = "html", caption = "Indices de modificación")
```

Indices de modificación

lhs	op	rhs	mi	epc
N21	~~	N36	44.13	0.27
E7	~~	E32	29.84	0.24
N36	~~	E22	25.02	0.21
N21	~~	E22	23.99	0.19
E32	~~	E37	22.49	0.20
E12	~~	E42	22.41	0.20
E27	~~	E32	22.22	0.21
N36	~~	E17	20.13	0.19

Plot SEM



Modelo reespecificado

```
# 1. Especificar un modelo
modell <- "Neuro =~ N1+N6+N11+N16+N31+N41
          Extra =~ E7+E12+E17+E32+E37+E42
          E7 ~~ E32
          E7 ~~ E37"

# 2. Estimamos un modelo
fit1 = lavaan::sem(modell,
                  data = big5,
                  estimator = "WLSMV",
                  mimic = "Mplus",
                  ordered = TRUE)
```

correlación latente

```
SEM1 <- inspect(fit1, what = "std")$psi %>% as_tibble() %>%  
  mutate(across(where(is.numeric), round, 2))  
  
knitr::kable(SEM1, format = "html")
```

Neuro	Extra
1.00	-0.89
-0.89	1.00

bondades de ajuste - Modelo reespecificado

```
bondad1=lavaan::fitMeasures(fit1, c("chisq.scaled", "df.scaled", "srmr",  
  as.data.frame() %>% rename("value" = "."))  
knitr::kable(bondad1, format = "html", caption = "Bondades de ajuste")
```

	value
chisq.scaled	223.58
df.scaled	51.00
srmr	0.06
wrmr	1.15
cfi.scaled	0.92
tli.scaled	0.89
rmsea.scaled	0.08

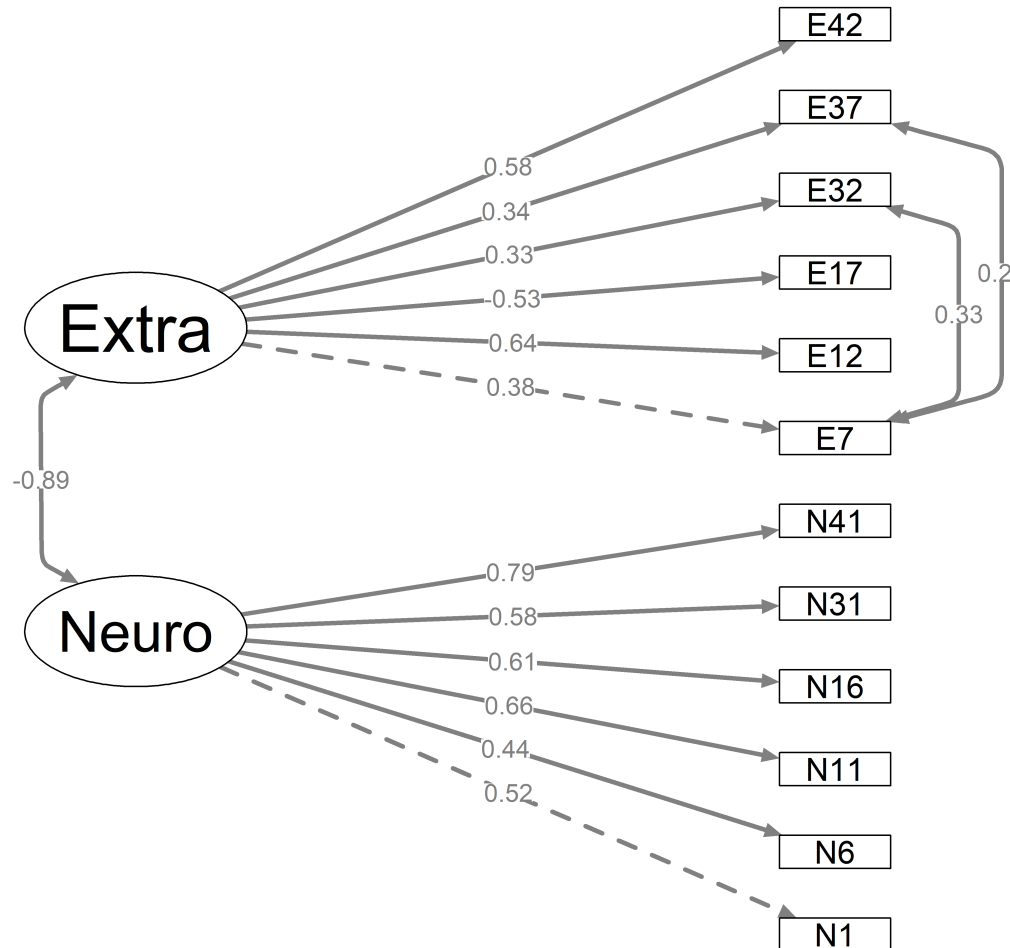
indices de modificación

```
mod1 <- modificationindices(fit1, sort. = T, minimum.value = 20)%>%  
  select(lhs:epc) %>% mutate(across(where(is.numeric), round, 2)) %>%  
  filter(op == "~~")  
knitr::kable(mod1, format = "html", caption = "Indices de modificaci
```

Indices de modificación 1

lhs	op	rhs	mi	epc
E32	~~	E37	42.23	0.27

Plot SEM Respecificado



Fiabilidad

```
fiabilidad <- semTools::reliability(fit1, what = c("omega3")) %>% round(2)
knitr::kable(fiabilidad, format = "html", booktabs = TRUE, valign = "top")
```

	Neuro	Extra
omega3	0.73	0.36

Limitaciones de SEM

1. Se requieren escalas cortas.
2. De preferencia unidimensionales.
3. Tamaños muestrales ($n > 300$)

Fuentes de referencia:

Sample Size Requirements of the Robust Weighted Least Squares Estimator

A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data

Gracias por su atención

Dr. José Luis Ventura León

Docente investigador

Web: joseventuraleon.com



Dr. José Ventura-León